http://mmg.tversu.ru ISSN 2311-1275

MMG *Mathematical Modelling and Geometry*

Volume 5, No 1, p. 21 – 29 (2017)

Optimization procedure for migration of DNA molecules

M. Meza Vargas[1,a] **and G. S. Sharov**[1,b]

[1] Faculty of Mathematics, Tver State University, Sadovyi per. 35, Tver, Russia
e-mail: [a] esmasanlame@gmail.com [b] Sharov.GS@tversu.ru

Received 28 January 2017, in final form 22 February. Published 24 February 2017.

Abstract. We investigate an optimization procedure for different patterns in biological experiments, in particular, the recent data for migration of DNA (deoxyribonucleic acid) molecules in electrophoresis. We describe the results of experiments, where the mobility of different DNA fragments in electric field was measured in two porous media: in polyacrylamide gel and free solution. We calculate the χ^2 function, that indicates the correspondence between the theoretical and experimental values of five free model parameters $D_1 \dots D_5$, which determine the mobility decrement of DNA fragments. In our optimization procedure we search the absolute minimum of χ^2 in the five-dimensional space of these free parameters D_i^{th} and obtain their optimal values for polyacrylamide gel and free solution. Finally we compare the results of our fitting procedure and the optimal values, used by the authors of the experiment.

Keywords: DNA, electrophoresis, mobility decrement

MSC numbers: 37N25, 65K10

1. Introduction

DNA (deoxyribonucleic acid) is the universal genetic code. The DNA molecule is a long double helix, comprising various chemical components — nucleotides. Each nucleotide is composed of a sugar molecule, a phosphate group and a nitrogenous base. There are four nitrogenous bases: adenine (A), guanine (G), thymine (T) and cytosine (C) [1].

In 1964, scientists began to investigate DNA by electrophoresis [2]. This is a method that separates macromolecules with different spatial configurations and electric charges. It plays an important role in the research of proteins and nucleic acids [3, 4]. In electrophoresis, DNA fragments are separated according to their size and electrical charge. These properties determine the mobility in the electric field [2, 3]. Applying an electric current allows DNA fragments to migrate from negative to positive electrodes. In this case, the velocity $\boldsymbol{v}$ of the molecule is proportional to the intensity of the electric field $\boldsymbol{E}$:

$$\boldsymbol{v} = \mu \boldsymbol{E}$$

The coefficient of proportionality μ in this relation is referred to as mobility.

Many authors [2, 5, 6, 7] observed electrophoresis of DNA fragments in porous media and investigated how the mobility of fragments in electric field depends on the size of the fragment, the order of nitrogenous bases, the composition of the medium, the gel concentration, the pore size and other parameters.

2. Mobility of curved DNA fragments

In Ref. [2], the mobility of different DNA fragments has been measured in the porous structure that contained free solution and polyacrylamide gel. The DNA fragments had various degrees of curvature in their structure. The authors have observed that the fragments with greater curvature migrated slowly in the polyacrylamide gel as well as in the free solution which contained 40 mM TAE buffer [40 mM Tris (2-amino-2-hydroxymethyl-propane-1,3-diol) and 1 mM EDTA (Ethylenediaminetetraacetic acid), brought to ph 8.0 with glacial acetic acid].

The authors used the DNA fragment taken from the VP1 gene in the SV40 minichromosome and determined the fragment's mobility μ according to the following formula:

$$\mu = \frac{\ell}{Et},$$

where ℓ is the distance traveled during the time t. The goal of this study was to understand the relationship between the DNA fragment mobility and its degree of curvature, associated with the presence or absence of certain tracts of nitrogenous bases in the fragment. The authors denoted these tracts as A_6, A_4T, A_4, A_3T_4, T_7 and discovered the bases in the DNA fragment with strong curvature that decrease mobility μ in comparison with the maximum fragment mobility μ_0 corresponding to the minimum curvature in the DNA fragment. The authors of Ref. [2] introduced the quantity called the mobility decrement:

$$D = 100 \cdot \frac{\mu_0 - \mu}{\mu_0}. \tag{1}$$

where μ is the mobility of a fragment. The results of Ref. [2], in particular, the experimental values of mobility decrement (1) in polyacrylamide gel (D^g) and in free solution (D^s) are shown in Table 1.

The first column in Table 1 corresponds to the DNA fragment taken from VP1 gene in the minichromosome. The parent 199P fragment includes all tracts of nitrogenous bases A_6, A_4T, A_4, A_3T_4, T_7. All of them were modified by site-directed mutagenesis using PCR (polymerase chain reaction), and 31 new fragments with all possible combinations of A-T tracts were generated. In the columns "D^g" and "D^s" of Table 1, the experimental data of mobility decrements (1) (D^g for gel and D^s for free solution) are shown. These values are compared with theoretical results ($D^g_{[2]}$ and $D^s_{[2]}$) from Ref. [2], as well as, with our own results (D^g_o and D^s_o), that were calculated in this paper using the optimization technique.

The theoretical mobility decrement for each type i of a fragment is calculated using the following formula [2]:

$$D_i^{th} = \sum_{j=1}^{5} a_{ij} D_j. \tag{2}$$

Here the structural coefficient a_{ij} is equal to 1 if the tract A or T is present in the fragment and 0 otherwise (see Table 1); D_1, D_2, D_3, D_4, D_5 are five free model parameters which determine the fitted mobility theoretical decrement D_i^{th} of the DNA fragment. The calculated D_i^{th} are equal to $D^g_{[2]}$ or $D^s_{[2]}$ in Table 1, if we substitute in Eq. (2) the optimal values $D_1, \ldots, D_5$ from Ref. [2]. The optimal values of D_j, calculated below, yield the values $D_i^{th} = D^g_{oi}$ and $D_i^{th} = D^s_{oi}$ in Table 1.

When any experiment is executed, the statistical errors exist [8]. Hence, there are technical errors (instrumental) and systematic errors (human), resulting in the differences between the values D^g, $D^g_{[2]}$ and D^s, $D^s_{[2]}$ in Table 1. Each free parameter $D_1 \ldots D_5$ is used to achieve the minimal difference between theoretical and experimental results. Accordingly, we choose these parameters and use them in the chi-square test

$$\chi^2 = \sum_{i=1}^{n} \frac{(D_i - D_i^{th})^2}{\sigma_i^2}, \tag{3}$$

to obtain the best results. Here D_i is the mobility decrement (1) obtained in the experiment and $D_i^{th} = D_i^{th}(D_1 \ldots D_5)$ is the theoretical value of this mobility decrement calculated using the formula (2). The authors of Ref. [2] did not show the errors σ_i for all measurements, so we assume $\sigma_i = 1$ for all measured values

Fragment	A and T present					D^g	$D^g_{[2]}$	D^g_o	D^s	$D^s_{[2]}$	D^s_o
199P	A_6	A_4T	A_4	A_3T_4	T_7	23.5	21.2	21.181	1.63	1.63	1.6481
199F	–	A_4T	A_4	A_3T_4	T_7	18.2	15.6	15.137	1.16	1.27	1.2541
199B	A_6	–	A_4	A_3T_4	T_7	16.6	14.4	14.55	1.30	1.15	1.1746
199C	A_6	A_4T	–	A_3T_4	T_7	22.0	20.7	21.012	1.51	1.47	1.5261
198A	A_6	A_4T	A_4	–	T_7	14.4	16.3	16.325	1.26	1.29	1.2695
199E	A_6	A_4T	A_4	A_3T_4	–	18.1	17.8	17.70	1.24	1.34	1.3681
199Q	–	–	A_4	A_3T_4	T_7	9.8	8.8	8.506	0.85	0.79	0.7806
199R	–	A_4T	–	A_3T_4	T_7	13.0	15.1	14.968	1.19	1.11	1.1321
198X	–	A_4T	A_4	–	T_7	7.4	10.7	10.281	1.06	0.93	0.8755
199G	–	A_4T	A_4	A_3T_4	–	11.4	12.2	11.656	0.93	0.98	0.9741
199D	A_6	–	–	A_3T_4	T_7	12.9	13.9	14.381	1.10	0.99	1.0526
198C	A_6	–	A_4	–	T_7	6.7	9.5	9.694	0.71	0.81	0.7960
199P	A_6	–	A_4	A_3T_4	–	10.6	11.0	11.069	1.02	0.86	0.8946
198B	A_6	A_4T	–	–	T_7	16.7	15.8	16.156	1.12	1.13	1.1475
199V	A_6	A_4T	–	A_3T_4	–	16.3	17.3	17.531	1.24	1.18	1.2461
199U	A_6	A_4T	A_4	–	–	12.6	12.9	12.844	0.77	1.00	0.9895
199S	–	–	–	A_3T_4	T_7	8.5	8.3	8.337	0.56	0.63	0.6586
198J	–	–	A_4	–	T_7	4.5	3.9	3.65	0.24	0.45	0.402
199L	–	–	A_4	A_3T_4	–	2.1	5.4	5.025	0.63	0.50	0.5006
198I	–	A_4T	–	–	T_7	10.0	10.2	10.112	0.77	0.77	0.7535
199H	–	A_4T	–	A_3T_4	T_7	11.2	11.487	11.531	0.82	0.82	0.8521
199O	–	A_4T	A_4	–	–	6.0	7.3	6.80	0.50	0.64	0.5955
198D	A_6	–	–	–	T_7	10.0	9.0	9.525	0.65	0.65	0.674
199T	A_6	–	–	A_3T_4	–	9.4	10.5	10.90	0.54	0.70	0.7726
198H	A_6	–	A_4	–	–	6.9	6.1	6.213	0.53	0.52	0.516
198G	A_6	A_4T	–	–	–	15.2	12.4	12.675	1.06	0.84	0.8675
198E	–	–	–	–	T_7	3.2	3.4	3.481	0.35	0.29	0.28
199I	–	–	–	A_3T_4	–	4.8	4.9	4.856	0.53	0.34	0.3786
199N	–	–	A_4	–	–	2.1	0.5	0.169	0.36	0.16	0.122
199J	–	A_4T	–	–	–	6.6	6.8	6.631	0.74	0.48	0.4735
199F	A_6	–	–	–	–	5.9	5.6	6.044	0.69	0.36	0.394
199K	–	–	–	–	–	1.8	0.0	0.000	0.50	0.00	0.00

Table 1: DNA fragments with their different A-T tracts and mobility decrements. The index "g" corresponds to polyacrylamide gel, "s" is used for free solution.

$D_i = D_i^g$ in the case of polyacrylamide gel, and $\sigma_i = 0.1$ for all free solution values D_i^s.

The chi-square test was proposed by Karl Pearson to achieve agreement between experimental data and any hypothesis or theory for Gaussian (standard normal) distributions [8, 9]. The normal distribution plays an important role in mathematical statistics and describes the selectivity of various functions in the distribution based on the observed results. This probability distribution is used to construct confidence intervals and statistical tests [8].

In our particular research, we calculate χ^2 applying the formula (3) to experimental and theoretical data, moreover; we obtain the value of χ^2 dependent upon D_1, D_2, D_3, D_4, D_5 for polyacrylamide gel and free solution. Afterwards, we search the minimum $\min\limits_{D_1-D_5} \chi^2$ for 5 free parameters D_i.

This investigation was done with the computational mathematical package MATLAB optimized for solving engineering and scientific problems [10]. At the beginning we create a matrix with 7 columns, where the first five columns contain a_{ij} (Table 1). In the sixth and seventh columns, we keep the experimental data D_i^g and D_i^s. We fix the limits of each free parameter D_j and change these limits if the point of minimum χ^2 appears to be beyond this range. In particular, for the polyacrylamide gel, the first parameter D_1 varied within $5.95 \leq D_1 \leq 6.1$ with $\Delta D_1 = 0.001$. The second free parameter D_2 varied in the segment $[6.59, 6.67]$ with $\Delta D_2 = 0.005$. These limits were modified in the process of numerical calculations.

To achieve the minimum of χ^2, we create 5 cycles, where we fix each parameter D_1, D_2, D_3, D_4, D_5 respectively and calculate χ^2 for each set of D_i . In the first (external) cycle, we vary the parameter D_5 in the range $D_5 \in [3.45, 3.52]$ with $\Delta D_5 = 0.001$.

For each given value of D_5 we pass all points in the rectangle in (D_1, D_2) plane with 2 free parameters D_3, D_4 remaining unchanged. For each point (pair D_1, D_2) we calculate the minimum value over D_3, D_4: $m_{3,4}(D_1, D_2, D_5) = \min\limits_{D_3,D_4} \chi^2$. This minimum is calculated over the rectangle (chosen preliminary) in (D_3, D_4) plane, and we control, that at each step a position of this minimum point to remain inside the chosen rectangle.

At the next step we search the minimum of the function $m_{3,4}(D_1, D_2, D_5)$ over parameters D_1 and D_2 (the rectangle in the plane (D_1, D_2)): $m_{1-4}(D_5) = \min\limits_{D_1,D_2} m_{3,4}(D_1, D_2, D_5)$.

The results of this procedure for polyacrylamide gel are shown in Fig. 1. The panel **(A)** shows isolevel lines of the function $m_{3,4} = \min\limits_{D_3,D_4} \chi^2$ in the plane (D_1, D_2) for the fixed value D_5, and **(B)** displays the χ^2 in the plane D_3, D_4 for optimal parameter values D_1 and D_2. These isolevel lines correspond to the values $m_{abs} + 0.001$, $m_{abs} + 0.003$, $m_{abs} + 0.01$, where m_{abs} is the absolute minimum of the χ^2 function (3).

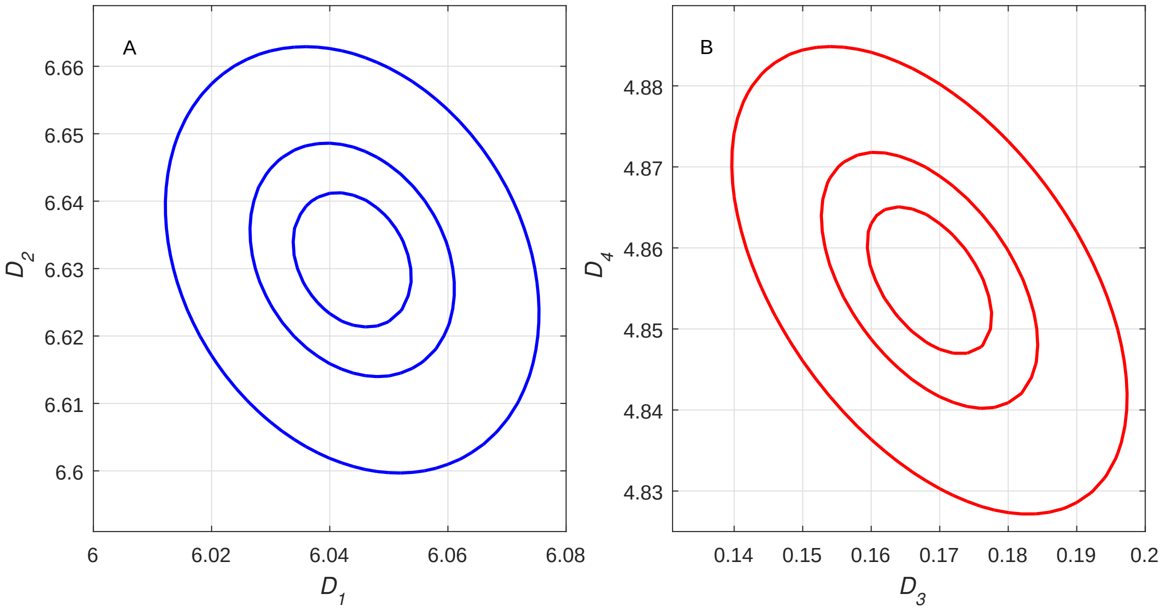

Figure 1: **(A)** Contour plots (isolevel lines) of $m_{3,4}(D_1, D_2, D_5)$ (gel) in the plane (D_1, D_2) for $D_5 = 3.481$; **(B)** Contour plots χ^2 in the plane (D_3, D_4) for optimal parameters $D_1 = 6.044$, $D_2 = 6.631$ and $D_5 = 3.481$.

When all calculations in the external cycle (for all values of the D_5) are performed, we obtain a graphic representation for the function $\min\limits_{D_1-D_4} \chi^2$ that depends on D_5. This function is shown in the left-hand panels of Fig. 2. The absolute minimum of χ^2 in this numerical experiment is $m_{abs} = \min\limits_{D_1 \ldots D_5} \chi^2 = 76.798$, this value is indicated in Table 2.

Parameter	D_1	D_2	D_3	D_4	D_5	χ^2
$D^g_{[2]}$	5.6	6.8	0.5	4.9	3.4	79.55
D^g_o	6.044	6.631	0.169	4.856	3.4816	76.798
$D^s_{[2]}$	0.36	0.48	0.16	0.34	0.29	84.66
D^s_o	0.394	0.4735	0.122	0.3786	0.28	80.83

Table 2: Optimal free parameters and χ^2 values for A-T tracts. The index g corresponds to polyacrylamide gel and s is used for free solution

We can see that this value is lower than the value $\chi^2 = 79.55$ obtained in the Ref. [2]. We performed analogous calculations for the free solution; the results are shown in Table 2 and in the right-hand panels of Fig. 2.

Fig. 2 presents the graphical analysis of the minimum values of χ^2 over 4 parameters D_i as functions of the fifth parameter when all calculations are completed

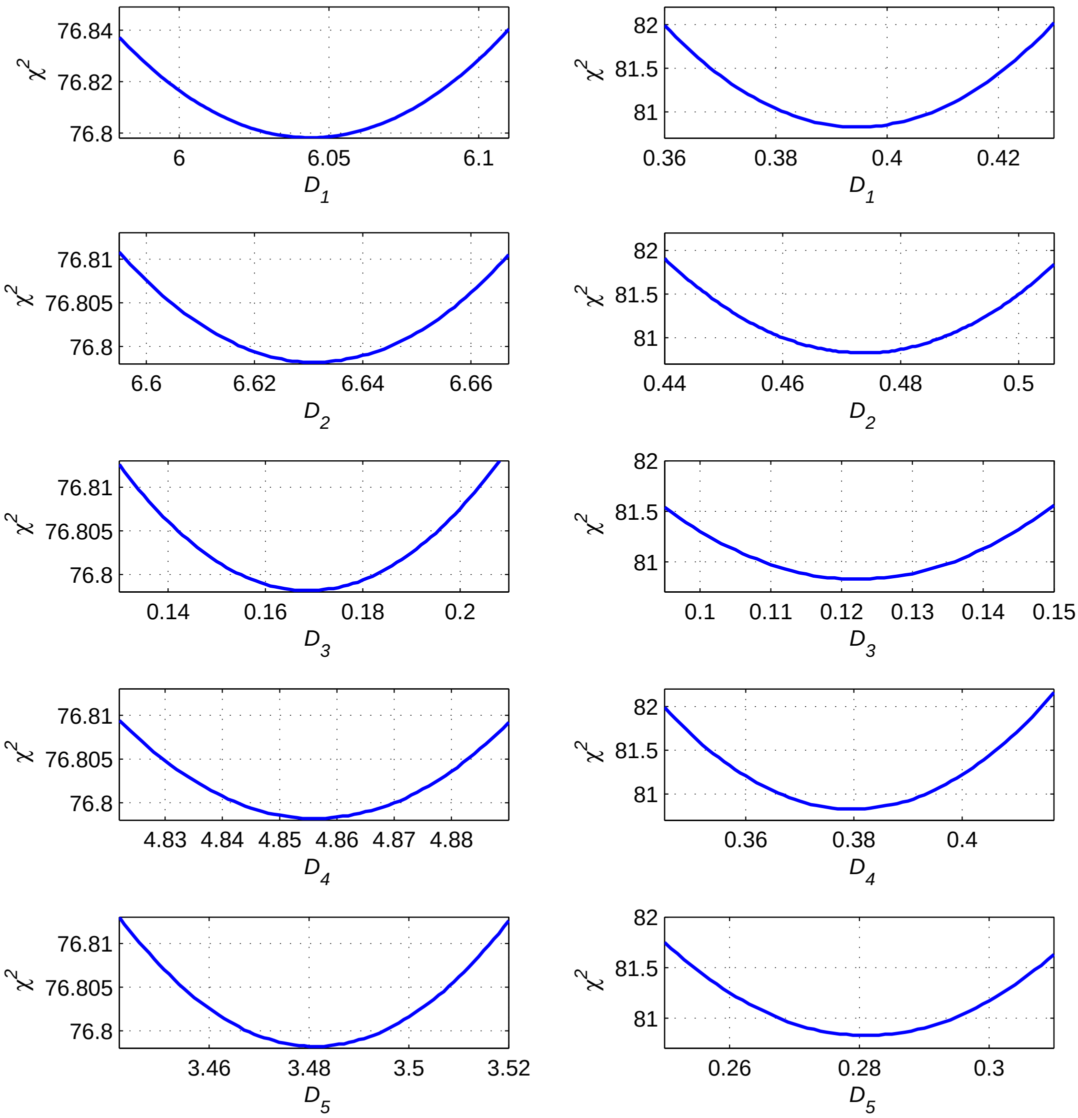

Figure 2: Plots of minimal functions $m(D_i) = \min\limits_{all\ D_j \neq D_i} \chi^2$ depending on $D_1, \ldots, D_5$ for polycrylamide gel (the left-hand panels) and for free solution (the right-hand panels)

in the external cycle. Basically, this procedure was described above. Here we can observe the function $m(D_i) = \min \chi^2$ depends on $D_1, \ldots, D_5$ to calculate the absolute minimum of χ^2 and to get each optimal value for 5 free parameters in this numerical experiment. The left-hand panels of Fig. 2 show all plots $m(D_i) = \min \chi^2$ for each parameter $(D_1, \ldots, D_5)$ in polyacrylamide gel and the right-hand panels display $m(D_i) = \min \chi^2$ in free solution for the same 5 free parameters.

3. Conclusion

This numerical experiment clearly indicates that as a result of our mathematical calculations D_o^g and D_o^s correspond better to the experimental values in comparison with the values, used by the authors of the experiment [2]. It means that the values $m_{abs}^g = \min\limits_{D_1 \dots D_5} \chi^2 = 76.798$ and $m_{abs}^s = \min\limits_{D_1 \dots D_5} \chi^2 = 80.83$ obtained in this paper are the absolute minima for χ^2 function for polyacrylamide gel and free solution. The differences $\Delta m = \min \chi^2\big|_{[2]} - \min \chi^2\big|_{our}$ between these minima and the corresponding minimal values in Ref. [2] are: $\Delta m^s = 3.83$ and $\Delta m^g = 2.752$. Our absolute minima were obtained in this numerical experiment as a result of the optimization procedure over each of 5 free parameters $D_1, \dots, D_5$; their optimal values are presented in Table 2.

References

[1] Stellwagen N. C. et al, The use of gel and capillary electrophoresis to investigate some of the fundamental physical properties of DNA, Electrophoresis (2002) **23** No. 2, 167 – 175.

[2] Stellwagen A. E., Lu Y. and Stellwagen N. C., Curved DNA molecules migrate anomalously slowly in free solution, Nucleic Acids Research (2005) **33** No. 14, 4425 – 4432.

[3] Haran T. E. et al, Characterists of migration patterns of DNA oligomers in gels and the relationship to the question of intrinsic DNA bending, Journal of the American Chemical Society (2004) **126** No. 8, 2372 – 2377.

[4] Merling A. et al, A-tract polarity dominate the curvature in flanking sequences, Biochemistry (2003) **42** No. 17, 4978 – 4984.

[5] Lu. Y. et al., Intrinsic curvature in the VP1 gene of SV40: comparison of solution and gel results, Biophysical (2005) **88** No. 2, 1191 – 1206.

[6] Stellwagen N. C., Curved DNA molecules migrate anomalously slowly in polyacrylamide gels even at zero gel concentration, Electrophoresis (2006) **27** No. 5, 1163 – 1168.

[7] Liao W. C. et al., Conformation dependence of DNA electrophoretic mobility in a converging channel, Electrophoresis (2010) **31** No. 16, 2813 – 2821.

[8] McHugh M. L., The Chi-square test of independence, Biochemia Medica (2013) **23** No. 2, 143 – 149.

[9] Gonzalez O., Li J., Modeling the sequence-dependent diffusion coefficients of short DNA molecules, The Journal of Chemical physics (2008) **129** 165105.

[10] Taylor A. et al, GeneNet toolbox for matlab: a flexible platform for the analysis of gene connectivity in biological networks, Bioinformatics (2015) **31** No. 3, 442 – 444.